

Tato práce popisuje problém extrakce relací mezi pojmenovanými entitami na úrovni vět, za předpokladu, že pojmenované entity jsou již v textu označeny, na doméně policejních zpráv napsaných protidrogovým oddělením Policie České Republiky.

Použili jsme různé metody strojového učení v kombinaci se funkcemi stromových kernelů a metodami založenými na pravidlech větné syntaxe.

Žádná z použitých metod nedosáhla uspokojivých výsledků na datech poskytnutých Policií České Republiky. Následující analýza odhalila, že v označování poskytnutých dat chybí mnoho relací, které jsou při čtení člověkem zřejmé. To se ukázalo jako důvod, proč metody strojového učení s učitelem neuspěly.

Dále v této práci představujeme několik pravidel pro určení relací, která jsme identifikovali ručně.

Poznatky v této práci mohou být nápomocné pro další výzkum a zpracování těchto policejních zpráv.